A Machine-Learning approach to the estimation of CERES TOA fluxes

Bijoy Vengasseril Thampi[1], Takmeng Wong[2]
Constantine Lukashin[2]

[1]Science Systems and Applications, Inc., Hampton, VA
[2]NASA Langley Research Center, Hampton, VA

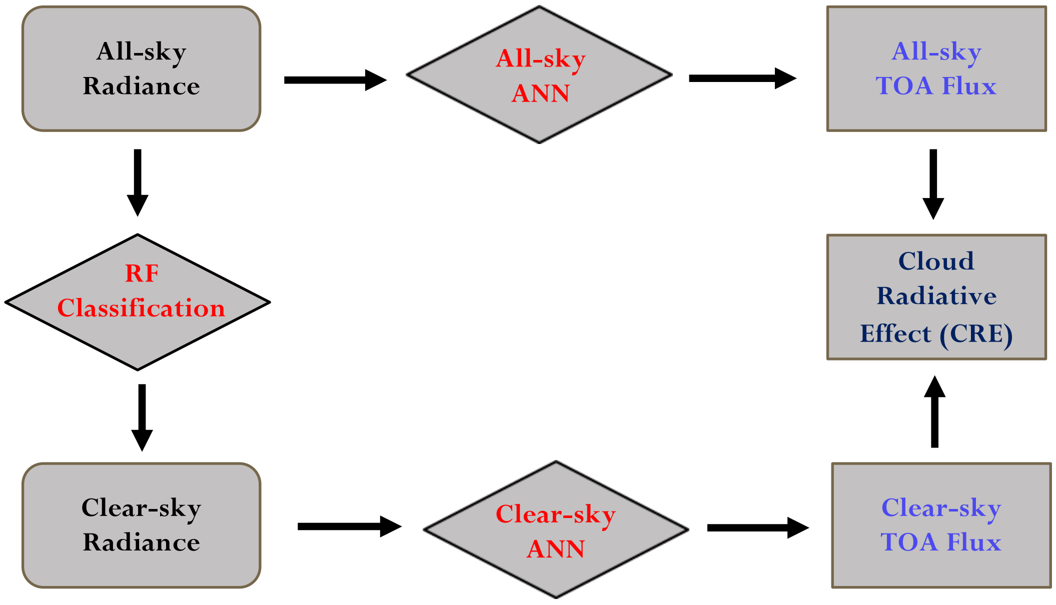
CERES Science Team Meeting, 28-30 April 2020

Objective

- To improve CERES stand-only/imager-independent TOA clear-sky and all-sky radiative fluxes using modern Machine Learning algorithms
- Current CERES stand-only TOA fluxes from CERES ERBE-like product are based on 30-year old legacy ERBE algorithms and are known to have larger uncertainty than CERES SSF TOA fluxes due to scene misclassification and ADM errors.
- These deficiencies are addressed using two sets of Machine Learning algorithms: Random Forests (RF) and Artificial Neural networks (ANN)

Methodology

- Scene Classification - **Random Forests (RF)** method
  - Developed by Breiman (2000)
  - Adopted for CERES –Thampi et al. (2017)
- TOA Flux estimation – **Artificial Neural network (ANN)** method
  - ANN methodology outlined in Lukashin and Loeb(2003)



Machine learning Algorithms

Random Forests (RF)	Artificial Neural networks (ANN)
▪ is an ensemble learning method for classification and regression. ▪ operate by constructing a multitude of decision trees and outputting the class that gets maximum number of votes from the forest. ▪ Main advantages are: • *faster runtimes* • *can deal with unbalanced and missing data* • *ability to handle data without preprocessing or rescaling.*	▪ based on a large number of neural units loosely modelling the way a biological brain solves problem. ▪ exceptionally good at performing pattern recognition and other tasks that are very difficult to program using conventional techniques. ▪ Programs that employ neural nets are also capable of learning on their own and adapting to changing conditions.

CERES ML Input

Input Variables	IGBP surface types
Solar & viewing zenith- angles	Water bodies
Relative azimuth angle	Bright Desert
CERES Shortwave (SW) and Longwave (LW) broadband radiances	Dark Desert
	Grasslands
LW surface emissivity	Croplands and cities
Broadband surface- albedo	Evergreen Forests
Surface skin temperature	Deciduous Forests
Precipitable water	Woody Savannas and Shrub lands
	Permanent and Fresh- snow
	Sea Ice

CERES ML output

Machine -Learning Output
TOA All-sky LW Flux
TOA All-sky SW Flux
TOA Clear-sky LW Flux
TOA Clear-sky SW Flux
TOA SW CRE
TOA LW CRE
TOA NET CRE

CERES Aqua SSF dataset is used as input data source

Training data : 2003 - 2014
Test data : 2015

Random Forests : Classification Accuracy

- Accuracy of RF scene classification is expressed using Misclassification rate (MCR in %) for each surface Type.
- In general, MCR values are higher for nighttime data than Day-time data. Also MCR values are higher for Snow and Sea ice surface while relatively lower for forests and crops

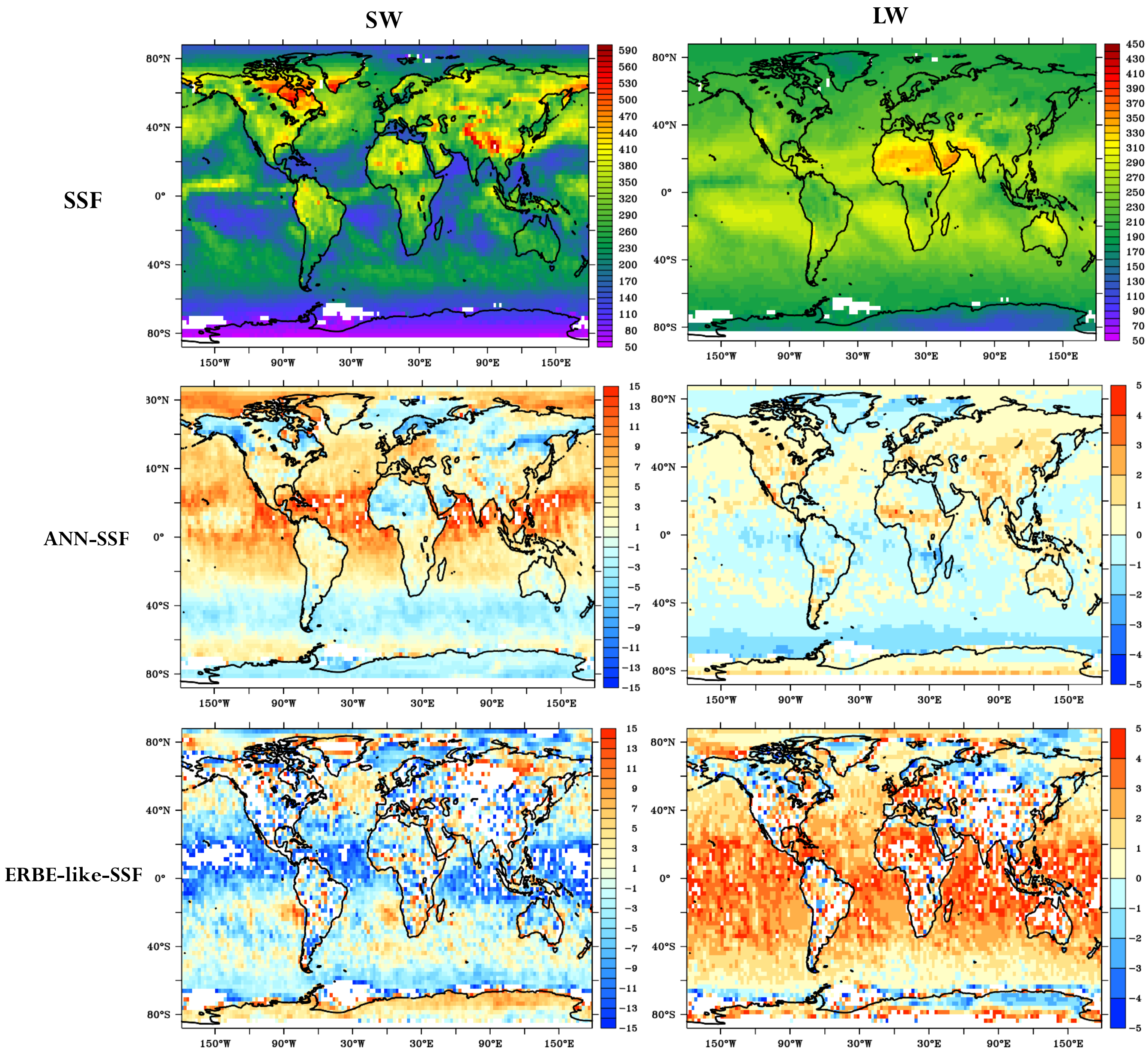
Surface Type	Day-time			Night-time		
	MCR (%)			MCR (%)		
	January	April	July	January	April	July
Everforests	1.1	0.6	2.0	9.0	6.3	16.6
Dcforests	8.6	3.5	3.6	24.3	12.8	20.3
Woodshrubs	12.4	1.5	3.2	24.8	12.6	19.1
Darkdeserts	12.1	9.7	7.8	15.7	14.9	12.4
Brightdeserts	16.8	13.2	11.4	13.6	19.7	12.7
Grass	10.2	7.5	7.1	25.7	22.5	17.8
Crops	6.6	4.2	3.7	23.8	14.0	15.8
Snow	25.2	24.8	19.6	31.9	39.4	37.7
Seaice	16.9	43.0	20.8	33.4	27.7	13.2
Waterbody	7.7	6.5	5.3	16.2	16.3	12.5

Percentage of Clear-sky Data Points
(SSF vs RF vs ERBE-like)

Month	Daytime (%)			Nighttime(%)		
	SSF (Truth)	RF	ERBE-like	SSF (Truth)	RF	ERBE-like
JAN	12.7	16.5	33.6	4.0	10.2	23.2
APR	8.0	14.6	28.6	7.5	14.0	15.7
JUL	6.8	10.5	27.3	7.4	13.0	24.8

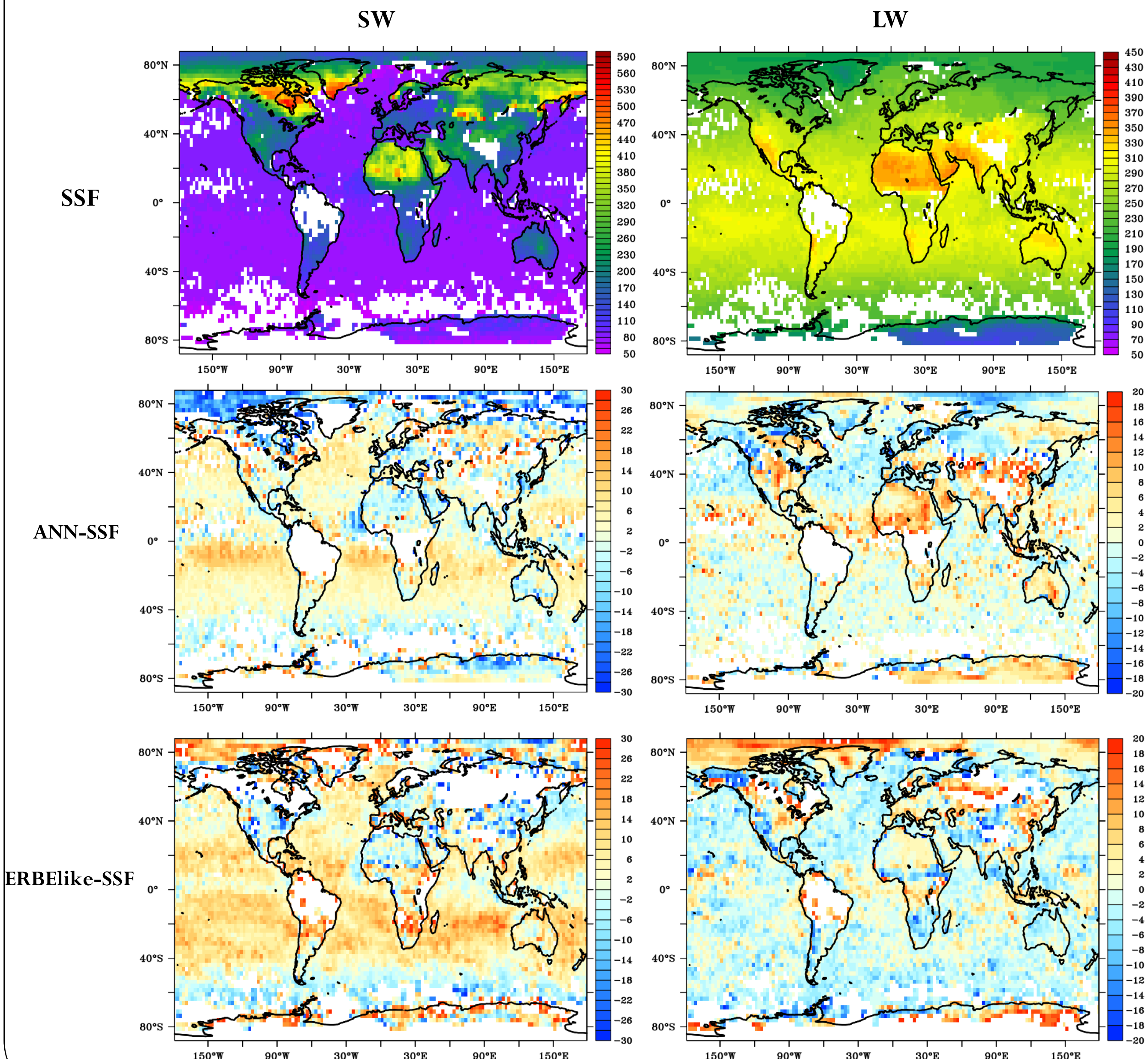
- Percentage of CERES SSF clear-sky data points (truth set) w.r.t. all-sky data is between 6.8 to 12.7% for daytime and 4.0 to 7.5% for nighttime.
- Random Forest derived clear-sky data points (<20%) aligns closer to the CERES SSF dataset compared to the ERBE-like dataset (>20%) for all months.

TOA Day-time All-sky Flux: *SSF vs ANN vs ERBE-like* (April 2015)



- Comparing to SSF data, ANN overestimates (**yellow**) the TOA daytime all-sky SW fluxes in the tropical oceans while underestimates (**blue**) the corresponding fluxes over land areas during April 2015.
- For LW fluxes, ANN overestimates (**yellow**) over land and underestimates (**blue**) over the oceans.
- The reverse is true for ERBE-like with large underestimation (**dark blue**) for SW and overestimation (**orange and red**) of LW flux
- ANN TOA daytime all-sky fluxes perform better than corresponding ERBE-like fluxes with smaller differences relative to SSF.

TOA Day-time Clear-sky Flux: *SSF vs ANN vs ERBE-like* (April 2015)



- Global mean map of TOA daytime clear-sky flux shows lower SW values over oceans.
- The difference between ANN and SSF derived TOA flux is higher over polar regions (i.e., snow and sea ice surface).
- In general, ANN TOA daytime clear-sky fluxes (SW and LW) perform better than the corresponding ERBE-like fluxes with smaller differences relative to SSF.

Seasonal month Comparisons (Aqua Overpass Time)

- In general, the ANN TOA SW fluxes (all-sky and clear-sky) are closer to the SSF values compared to ERBE-like.
- ANN minus SSF clear-sky SW difference are < -1.5 Wm^{-2} while that for ERBE-like minus SSF is usually > 1.5 Wm^{-2}.
- For LW day and night TOA all-sky fluxes, the difference is lower for ANN (< 1 Wm^{-2}) compared to those of ERBE-like (> 1 Wm^{-2}).
- RMSD values (in bracket) also show lower values for SW and LW for ANN compared to ERBE-like.

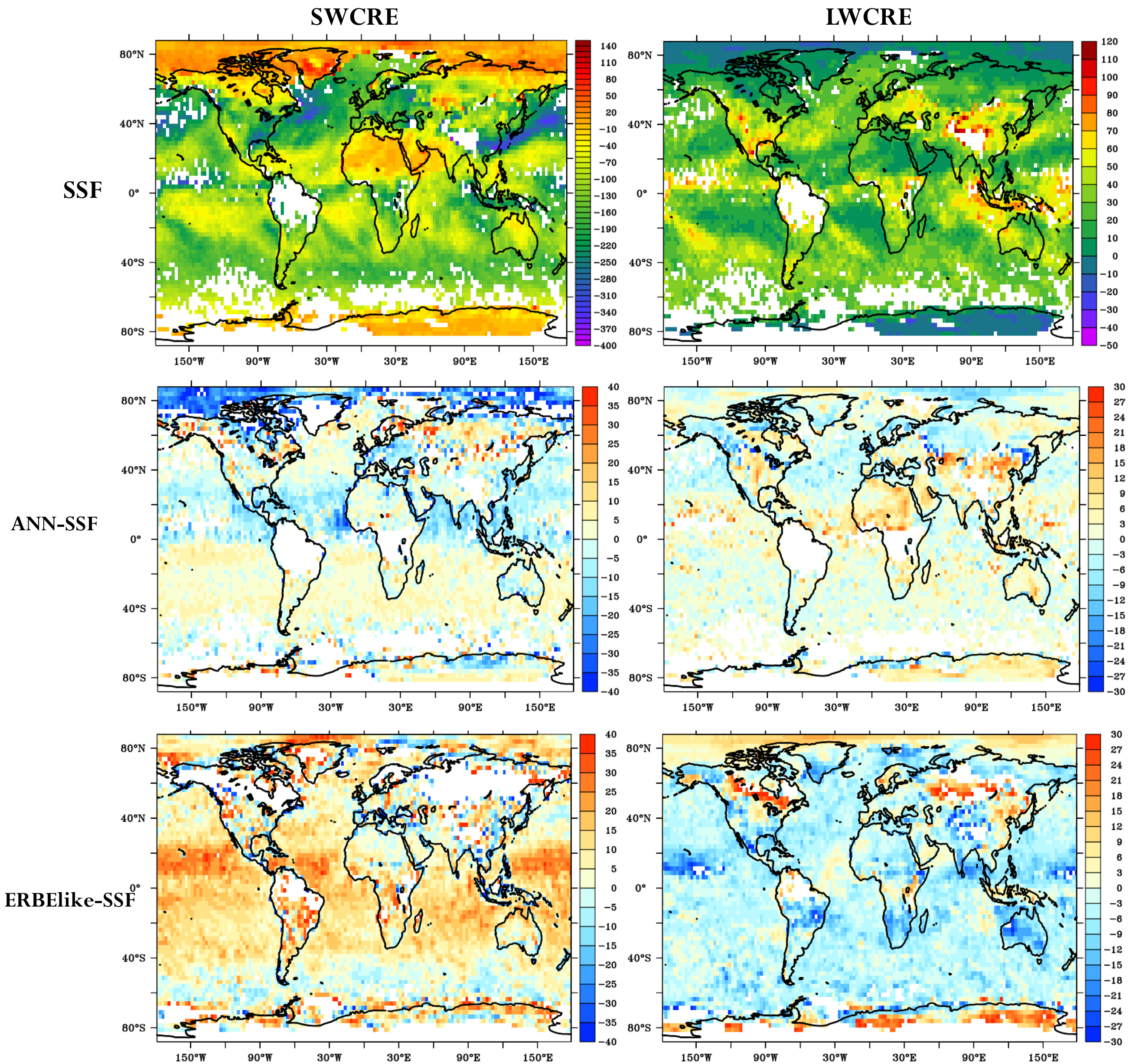
TOA Flux	Month	All- Sky			Clear-sky		
		SSF	ANN - SSF (RMSD)	ERBE-like - SSF (RMSD)	SSF	ANN - SSF (RMSD)	ERBE-like - SSF (RMSD)
SW	JAN	248.3	2.7 (20.6)	-2.2 (29.0)	114.2	-1.1 (30.2)	3.3 (37.9)
	APR	237.6	2.5 (18.8)	-3.8 (26.8)	119.9	-0.2 (26.4)	1.0 (39.4)
	JUL	221.7	0.7 (23.4)	-2.3 (27.6)	104.6	-0.7 (18.0)	5.0 (20.1)
LWDY	JAN	243.3	0.2 (6.1)	2.5 (8.0)	273.1	1.6 (5.2)	0.1 (6.8)
	APR	242.7	0.2 (6.3)	2.8 (8.4)	274.4	1.3 (6.2)	0.2 (7.7)
	JUL	251.5	-0.2 (6.7)	3.2 (8.7)	285.4	1.4 (5.4)	-3.4 (5.4)
LWNT	JAN	235.7	0.0 (1.6)	2.4 (3.6)	265.8	-0.1 (6.3)	5.6 (9.7)
	APR	235.1	0.2 (1.4)	2.6 (4.1)	265.7	-0.6 (6.4)	7.2 (13.6)
	JUL	241.4	0.0 (2.1)	2.5 (4.8)	271.4	0.6 (5.5)	3.2 (9.1)

2015 Annual Mean Comparisons (Aqua Overpass Time)

TOA Flux	All-sky			Clear-sky		
	SSF	ANN - SSF	ERBE-like - SSF	SSF	ANN - SSF	ERBE-like - SSF
SW	236.0	2.7	-3.2	111.5	-0.9	3.9
LWDY	245.2	-0.1	2.9	276.7	1.3	-0.7
LWNT	237.2	-0.2	2.6	267.8	-0.2	5.6

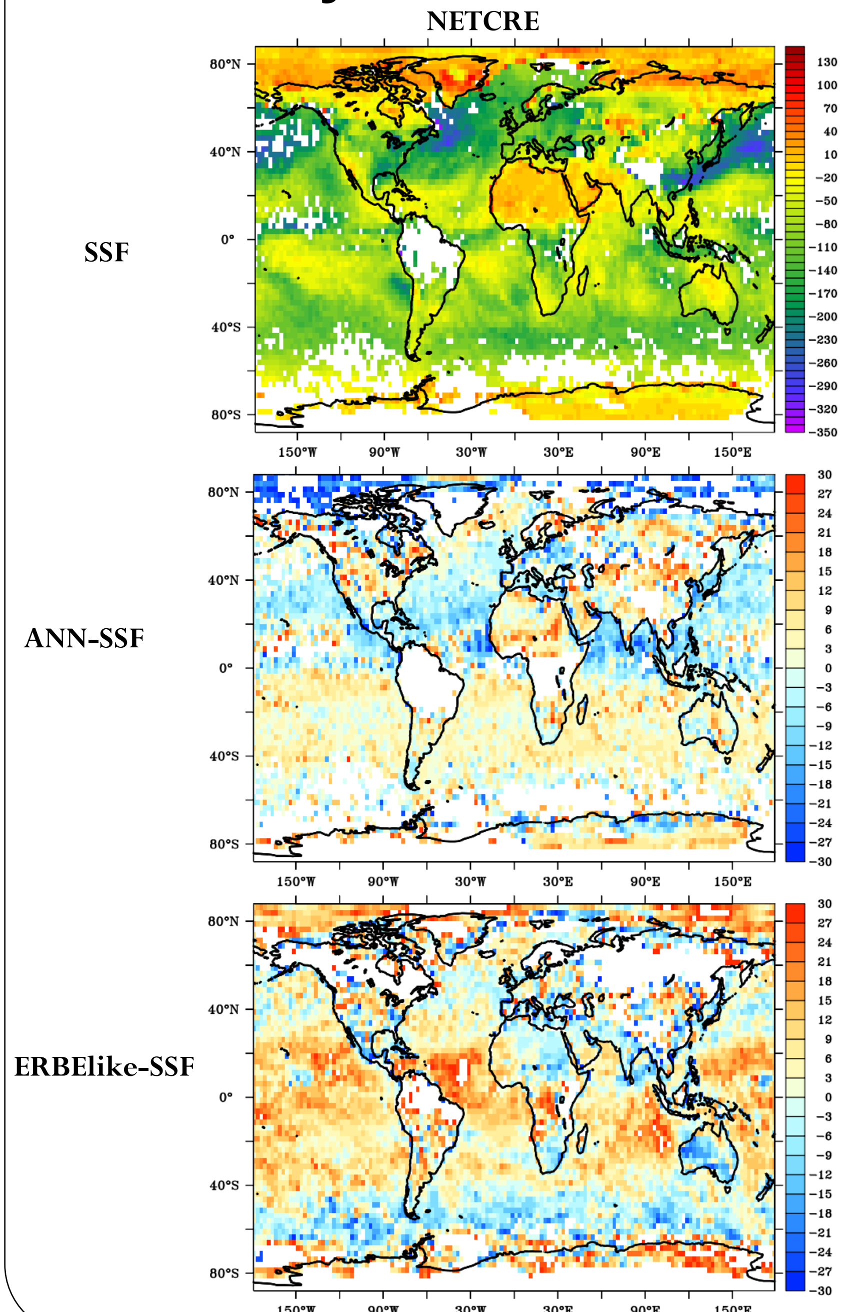
- ANN annual mean TOA fluxes are closer to the corresponding SSF values than those of the ERBE-like TOA fluxes.
- Mean SW difference between ANN and SSF is 2.7 (all-sky) and -0.9 Wm^{-2} (clear-sky) while it is > 3 Wm^{-2} for ERBE-like.
- Similar smaller LW mean differences (< -0.2 Wm^{-2}) are observed between ANN and SSF fluxes fluxes except for clear-sky daytime LW fluxes.

TOA Day-time CRE: *SSF vs ANN vs ERBE-like* (April 2015)



- Monthly mean Cloud radiative effect (CRE) values estimated using SSF and ANN dataset compares very well.
- **Bluish color represent negative values and orange-red color represent positive values**
- In general, SW and LW CRE difference is greater for ERBElike dataset compared to ANN dataset, indicating that the ***ANN CRE regional values are closer to SSF compared to ERBElike***
- CRE difference for ANN is negative over the polar regions, while it is positive for the ERBElike map indicating respective under and overestimation for the two dataset.

TOA Day-time NETCRE: *SSF vs ANN vs ERBE-like* (April 2015)



- **orange-red represent positive values and Bluish color color represent negative values**
- Similar to SW&LW CRE difference, NETCRE difference is also greater for ERBE-like dataset compared to ANN dataset.
- For global LWCRE and NETCRE values, both ANN and SSF shows almost similar values over the globe except for the polar regions where ANN is slightly underestimating while ERBE-like is overestimating over most of the regions

Mean Daytime TOA CRE (Aqua Overpass Time)

CRE	Month	SSF	ANN-SSF	ERBElike-SSF
SWCRE	January	-134.1	-3.8	5.4
	April	-117.7	-2.8	4.7
	July	-117.1	-1.4	7.3
	Annual	**-124.5**	**-3.6**	**7.2**
LWCRE	January	29.9	1.5	-2.4
	April	31.7	1.2	-2.6
	July	33.9	1.6	-6.6
	Annual	**31.5**	**1.4**	**-3.5**
NETCRE	January	-104.3	-2.3	3.0
	April	-85.9	-1.6	2.1
	July	-83.2	0.1	0.8
	Annual	**-93.0**	**-2.2**	**3.6**

- Monthly mean daytime TOA CRE values are shown below for the three seasonal months.
- In general ANN CREs are closer to the SSF compared to the ERBElike dataset for the three months shown.
- **ANN derived Annual CRE values also show better match with SSF compared to the ERBE-like dataset.**
- ANN SWCRE difference is usually lower than -4 Wm^{-2} while it is > 4 Wm^{-2} for the ERBElike
- Similarly, the ANN LWCRE difference is usually < -2 Wm^{-2} while it is > -2 Wm^{-2} for ERBElike

Summary

- A machine Learning methodology was developed to estimate the TOA fluxes from CERES TOA radiances without using Imager radiance.
- This ML methodology involving RF and ANN will be an improvement over the current CERES stand-alone ERBE-like method.
- RF method was able to identify clear-sky footprint much better than ERBE-like algorithms.
- Global mean TOA Fluxes and CRE estimated using the ANN method generally performs better compared to ERBE-like.
- Manuscript detailing the above results are in preparation for submission to *JAOT*.
- The new CERES ML TOA flux algorithm (version 1) is ready for implementation into the CERES data production system.

Thank you…